Economics Department

Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model Based Procedure

MASSIMILIANO MARCELLINO

ECO No. 97/30

EUI WORKING PAPERS

EUROPEAN UNIVERSITY INSTITUTE

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 97/30

Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model Based Procedure

MASSIMILIANO MARCELLINO

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in December 1997
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model Based Procedure

Massimiliano Marcellino*†
European University Institute

November 1997

Abstract

We suggest a simple non model based procedure to recover a time series from its temporally aggregated realizations. If additional assumptions on the underlying process are introduced, it is shown that the procedure is related to many of the former proposals in the literature. It can also be easily modified to deal with the estimation of missing observations and outliers, and with forecasting. Some important identification issues are finally discussed.

*I would like to thank Giampiero Gallo, Clive Granger, Soren Johansen, Marco Lippi, Grayham Mizon, Pravin Trivedi and participants at the 1996 Econometric Society European Meeting for useful comments. I alone am responsible for remaining errors.

†*Correspondence to:* Massimiliano Marcellino, Department of Economics, European University Institute, I-50016, San Domenico, Firenze, Italy.

1 Introduction

There is quite often a mismatch between the frequency of data generation and that of data collection, the latter being usually rather lower than the former. This problem is particularly relevant in economics where, in general, at most monthly data are available for key variables such as GNP or unemployment. Moreover, only recently monthly data collection has systematically started and, usually, only annual data are available up to the '50s and quarterly data up to the '70s. Hence, whenever the original realizations are of interest, they have to be recovered from the available temporally aggregated time series.

Thus, it is not surprising that the issue of temporal disaggregation of aggregated data has been extensively studied. At least five main approaches to the problem can be distinguished. The first method consists in recovering the disaggregated values by means of partial weighted averages of the aggregated ones, see e.g. Lisman and Sandee (1964). In the second method the disaggregated values are those which minimise a loss function under a compatibility constraint with aggregated data, see e.g. Boot *et al.* (1967), Cohen *et al.* (1971), Stram and Wei (1986). In the third method a similar problem is solved but it is assumed that a preliminary disaggregated series is available, so that the issue is how to best revise it in order for it to be compatible with aggregated data, see e.g. Denton (1971), Chow and Lin (1971), Fernandez (1981), and Litterman (1983). In the fourth method the assumption of a disaggregated ARIMA process is made, and an optimization problem similar to those in the second and third methods is solved, see e.g., respectively, Wei and Stram (1990) and Guerrero (1990). In the fifth method, the hypothesis of an ARIMA process is also maintained, but the disaggregated values are considered as missing observations, and their best estimators as such can be obtained, see e.g. Harvey and Pierse (1984), Kohn and Ansley (1986), Nijman and Palm (1986), and Gomez and Maravall (1994).

Our procedure belongs to the second method and assumes the mean squared disaggregation error, MSDE, as the loss function. This seems a reasonable choice because the MSDE is a widespread measure of the "goodness" of an estimator of the disaggregated data, and it is usually adopted to compare different procedures, see e.g. Chan (1993). In Section 2 we derive the linear estimator of the disaggregated data which minimizes this loss function and satisfies the aggregated data compatibility constraint. A modified procedure which keeps into account uncertainty about the disaggregated process is also introduced.

We make rather weak assumptions about the disaggregated process in solving the former problem. Actually, we only require the existence of moments up to order two. We then show that if some further assumptions are introduced, our proposal is related to other estimators in the second and fourth methods.

Some further issues such as the estimation of the unobservable series at different disaggregated frequencies, or with series at different aggregated frequency, the revision of preliminary estimates, and the derivation of a "disaggregator" which minimizes the generalized variance of the disaggregation errors are then discussed in Section 3.

In Section 4 we show that the procedure can be also applied to derive linear minimum MSDE estimators of missing observations and outliers, and to forecasting. We then relate our results to former findings in the literature on this topic, e.g., Whittle (1963), Chang *et al.* (1988).

In most of the aforementioned studies a substantial knowledge of the characteristics of the disaggregated process is assumed. Yet, usually, only aggregate information is available, and it is not sufficient to exactly identify the disaggregated characteristics of interest. In particular, to implement our procedure, as well as most of the other ones in the second and third methods, the variance covariance matrix of the disaggregated process is required. In Section 5 we show that it cannot be uniquely determined from that of the aggregated process, unless particular a priori restrictions are imposed. We suggest some conditions which are either only necessary or also sufficient for exact identification.

The identification problem is also present in the fourth and fifth methods for temporal disaggregation, because an infinite number of ARIMA processes are compatible with an aggregated one. The class of all admissible disaggregated ARIMA processes is derived in Section 6, a simple but rather stringent sufficient condition for the exact identification of one of them is proposed, while more general conditions are derived by Wei and Stram (1990). A simple example is then discussed to illustrate these issues. When there is uncertainty about the proper a priori restrictions to impose, a probabilistic statement about them, or about their implied variance matrices, can be made, and the modified procedure in Section 2 adopted.

Concluding remarks and a summary of the main results of the paper are proposed in Section 7.

2 Estimation of disaggregated values

Let us consider the zero mean stochastic process $x = \{x_t\}_{t=0}^{\infty}$ and assume that its realizations are subject to temporal aggregation before being observed. Hence, the observed values can be thought of as realizations of the process $y = \{y_\tau\}_{\tau=0}^{\infty} = \{\omega(L)x_{tk}\}_{t=1}^{\infty}$, where k indicates the frequency of aggregation, L is the lag operator, and $\omega(L) = \omega_0+\omega_1 L+...+\omega_{k-1}L^{k-1}$ characterizes the aggregation scheme. For example, $\omega(L) = 1+L+...+L^{k-1}$ in the case of flow variables and $\omega(L) = 1$ for stock variables.[1]

Let us also group the first N and Nk elements of y and x in the $N \times 1$ and $Nk \times 1$ vectors Y and X, and construct the $N \times Nk$ matrix W, with

$$W = \begin{pmatrix} \omega_0, \omega_1, ...\omega_{k-1} & 0, 0, ..., 0 & ... & 0, 0, ..., 0 \\ 0, 0, ..., 0 & \omega_0, \omega_1, ...\omega_{k-1} & ... & 0, 0, ..., 0 \\ ... & & & \\ 0, 0, ..., 0 & 0, 0, ..., 0 & & \omega_0, \omega_1, ...\omega_{k-1} \end{pmatrix},$$

so that $Y = WX$.

We assume that the expected loss due to a mismatch between X and its estimator $\widehat{X}$ corresponds to the mean squared disaggregation error

[1]The following analysis remains valid even if k or the weights in $\omega(L)$ change over time.

(MSDE), so that the problem that we wish to solve can be formulated as:

$$\min_{\widehat{X}} tr(E(X-\widehat{X})(X-\widehat{X})') \tag{1}$$

$$\text{s.t. } Y = WX. \tag{2}$$

Under the additional hypothesis that $\widehat{X}$ is a linear function of Y, $\widehat{X} = PY$, the solution is given in the following Proposition.

Proposition 1. The linear minimum MSDE estimator of X is

$$\widehat{X}^* = P^*Y = V_X W' V_Y^{-1} Y, \tag{3}$$

where V_X and V_Y are the variance covariance matrices of X and Y. Moreover, $\widehat{X}^*$ is unbiased, satisfies $Y = W\widehat{X}^*$, and it is $E(X-\widehat{X}^*)(X-\widehat{X}^*)' = V_X - V_X W' V_Y^{-1} W V_X$.

Proof Let us consider a generic linear estimator $\widehat{X} = PY = PWX$. The objective function can then be written as

$$tr(E(I-PW)XX'(I-PW)') = tr((I-PW)V_X(I-PW)').$$

For it to be minimised it is necessary that P^* satisfies the first order conditions

$$-V_X W' + P^* W V_X W' = 0.$$

This is also sufficient because the second order conditions are also satisfied, given that WV_XW' is a positive definite matrix. Furthermore, from (2) it follows that $WV_XW' = V_Y$. Hence, the linear minimum MSDE estimator is

$$\widehat{X}^* = P^*Y = V_X W' V_Y^{-1} Y.$$

Moreover,

$$E(\widehat{X}^*) = P^*E(Y) = P^*WE(X) = 0, \text{ and } W\widehat{X}^* = WV_XW'V_Y^{-1}Y = Y.$$

Finally,

$$
\begin{gathered}
E(X-\widehat{X}^*)(X-\widehat{X}^*)' = E(X-P^*WX)(X-P^*WX)' = \\
E((I-P^*W)XX'(I-P^*W)') = (I-P^*W)V_X(I-P^*W)' = \\
((I-P^*W)V_X)' - ((I-P^*W)V_XW'P^{*\prime})' = \\
(V_X - V_XW'V_Y^{-1}WV_X)' - 0 = \\
V_X - V_XW'V_Y^{-1}WV_X. \blacksquare
\end{gathered}
$$

$\widehat{X}^*$ can be interpreted as the projection of X on the space spanned by Y. When there are no disaggregation problems, i.e. $W = I$, the former Proposition implies that it is $\widehat{X}^* = X$ and $E(X-\widehat{X}^*)(X-\widehat{X}^*)' = 0$.

Notice that we have made no hypotheses at all about the stationarity of x or the functional form of its generating mechanism, except for the existence of the second order moments V_X. Actually, we have also required x to have zero mean, but this assumption can be simply relaxed by de-meaning x and y, and substituting them with the resulting processes. The non zero mean of x can then be simply recovered from that of y.[2]

Instead, if we assume that x is a weakly stationary Gaussian process, $\widehat{X}^*$ is the minimum MSDE estimator of X, because it coincides with $E(X|Y)$. In this case, $\widehat{X}^*$ also coincides with the solution of the problem

$$
\min_X X'V_X^{-1}X \tag{4}
$$

$$
\text{s.t. } Y = WX.
$$

Such a solution was provided for this general formulation of the problem but with $\omega(L) = 1 + L + ... + L^{k-1}$ by Stram and Wei (1986), and for particular values of V_X by Boot *et al.* (1967) and Cohen *et al.* (1971). It can be simply shown that the choice of $\omega(L)$ does not alter the form of the solution.

[2]When x has a time varying mean, an identification problem can arise. Actually, from $m_y = E(Y) = WE(X) = Wm_x$, if all the elements of m_x are equal (or at least k by k equal), they can be uniquely recovered from those of m_y. Otherwise, an infinite number of m_x are compatible with m_y, as can be easily derived if we interpret the condition $m_y = Wm_x$ as a set of N equations in Nk unknowns.

If instead the hypothesis that x is an ARMA process is maintained, $\widehat{X}^*$ coincides with the pure ARMA based minimum MSDE estimator in Guerrero (1990).

For Proposition 1 to be applicable, it is necessary to assume that x follows a particular process, or at least to specify a V_X. This hypothesis is often made in the literature, see e.g. the aforementioned studies (with the exception of Guerrero (1990)), or Harvey and Pierse (1984), Kohn and Ansley (1986), Nijman and Palm (1986), and Gomez and Maravall (1995) in the missing observations approach. Yet, it seems rather restrictive because many processes or V_Xs can be compatible with their aggregated counterparts, as we will see in more details in Sections 5 and 6. Hence, an alternative more satisfactory route can be the specification of a set of potential disaggregated processes, together with a *subjective* probabilistic statement about their likelihood.

To this end, we indicate with $\mathcal{M}$ the set of processes, $\mathcal{M} = \{M_i, i = 1, ..., m\}$, and with p_i the probability of M_i, where $\sum_{i=1}^m p_i = 1$. Notice that each process in $\mathcal{M}$ has only to be such that $WV_X^iW' = V_Y$. Thus, the problem becomes:

$$\min_{\widehat{X}} tr(E_M(E(X - \widehat{X})(X - \widehat{X})')) \tag{5}$$

$$\text{s.t. } Y = WX \ ,$$

and the solution is given in the following Proposition.

Proposition 2. The linear minimum expected MSDE estimator of X is

$$\widehat{X}_M^* = \overline{V}_X W' V_Y^{-1} Y = \sum_{i=1}^{m} \widehat{X}_{M_i}^* p_i, \tag{6}$$

where $\overline{V}_X = \sum_{i=1}^m V_X^i p_i$. Furthermore, it is unbiased and satisfies $Y = W\widehat{X}_M^*$.

Proof Let us consider again a generic linear estimator $\widehat{X} = PY = PWX$ and write the objective function as

$$\sum_{i=1}^{m} p_i tr((I - PW)V_X^i(I - PW)').$$

P^* has to satisfy the first order conditions

$$-\overline{V}_X W' + P^* W \overline{V}_X W' = 0,$$

while the second order conditions are satisfied because $W\overline{V}_X W' = \sum_{i=1}^m p_i W V_X^i W' = \sum_{i=1}^m p_i V_Y = V_Y$ is positive definite. Hence,

$$\widehat{X}_M^* = P^* Y = \overline{V}_X W' V_Y^{-1} Y$$

is the best linear estimator of X in the MSDE sense.

It is also

$$E(\widehat{X}_M^*) = P^* E(Y) = 0$$

and

$$W\widehat{X}_M^* = W\overline{V}_X W' V_Y^{-1} Y = \sum_{i=1}^m p_i W V_X^i W' V_Y^{-1} Y = Y. \blacksquare$$

Hence, $\widehat{X}_M^*$ is a convex linear combination of all the potential "disaggregators", whose weights depend on the relative probability of their associated disaggregated processes (or variance matrices). A continuous cumulative density function for the processes in $\mathcal{M}$, $F(M)$, could also be adopted. Under the assumption of existence of $\overline{V}_X = \int_{\mathcal{M}} V_M dF(M)$, the former formula for $\widehat{X}_M^*$ is still valid.

The procedure can be easily adapted to deal with the multivariate case. For example, in the bivariate case, we can partition Y and X into $(Y_1, Y_2)'$ and $(X_1, X_2)'$, where Y_1, Y_2, X_1 and X_2 are respectively $N_1 \times 1$, $N_2 \times 1$, $N_1 k \times 1$, and $N_2 k \times 1$ vectors of elements of the aggregated and disaggregated processes $y = (y_1, y_2)'$, $x = (x_1, x_2)'$, and $N_1 = N_2 = N/2$. Then, the optimal estimators of the disaggregated values are $\widehat{X}^* = (\widehat{X}_1^*, \widehat{X}_2^*)'$ in (3).

The procedure can be also used to recover disaggregated values from *linearly* aggregated series by simply redefining the involved variables. For example, X can be interpreted as a vector containing N variables for k agents, while in Y the variables are aggregated over agents.

Finally, joint temporal and linear disaggregation can be easily dealt with in a similar manner. This case is considered by Rossi (1982), whose method can be seen as a mixture of those by Denton (1971) and Chow and Lin (1971), which are reviewed in next Section, and in Di Fonzo (1990), who extends Chow and Lin's approach.

3 Some Extensions

We now extend the basic results and procedures in the former Section. First, we deal with estimation of the unobservable series at several disaggregation frequencies, and with observable series at different levels of aggregation. Then we consider the case where a preliminary estimate of X is available. Finally, we derive the optimal disaggregator when the objective function is the generalized variance of the disaggregation errors.

3.1 Multiple temporal frequencies

For simplicity but without loss of generality, let us assume that probability one is assigned to one process in $\mathcal{M}$, and that neither X_1 nor X can be observed where, in an obvious notation,

$$\begin{aligned} \underset{Nk_1\times 1}{X_1} &= \underset{Nk_1\times Nk}{W_2} \underset{Nk\times 1}{X}, \\ \underset{N\times 1}{Y} &= \underset{N\times Nk_1}{W_1} \underset{Nk_1\times 1}{X_1} = W_1 W_2 X = WX, \end{aligned} \tag{7}$$

with $k = k_1 k_2$.

To find the linear minimum MSDE estimators of X and X_1 we can apply the formula in (3), which yields $\widehat{X}^* = V_X W' V_Y^{-1} Y$ and $\widehat{X_1}^* = V_{X_1} W_1' V_Y^{-1} Y$. As an alternative, we can exploit $\widehat{X}^*$ or $\widehat{X}_1^*$ to derive, respectively, the best estimators of X_1 and X. For example, we can estimate monthly and quarterly data from annual data, or we can use the estimated monthly data to derive quarterly data, or vice versa. To this end, we have

Proposition 3. $\widehat{X}_1^* = W_2 \widehat{X}^*$ and $\widehat{X}^* = V_X W' (W_1 V_{\widehat{X}_1^*} W_1')^{-1} W_1 \widehat{X}_1^*$

.

Proof From (3) we know that $\widehat{X}_1^* = V_{X_1} W_1' V_Y^{-1} Y$ and $\widehat{X}^* = V_X W' V_Y^{-1} Y$. It follows that

$$W_2 \widehat{X}^* = W_2 V_X W_2' W_1' V_Y^{-1} Y = V_{X_1} W_1' V_Y^{-1} Y = \widehat{X}_1^*.$$

Moreover,

$$\begin{aligned} V_X W' (W_1 V_{\widehat{X}_1^*} W_1')^{-1} W_1 \widehat{X}_1^* &= V_X W' V_Y^{-1} W_1 V_{X_1} W_1' V_Y^{-1} Y = \\ &= V_X W' V_Y^{-1} Y = \widehat{X}^*. \blacksquare \end{aligned}$$

Hence, the best linear estimators of the aggregated series can be obtained by aggregating the most disaggregated estimators. Instead, there are no major computational advantages in using the aggregate estimators to determine the disaggregated ones.

A further topic to be considered is the role of observable series at different levels of aggregation. For example, we can assume that both Y and X_1 are observable, and we have to decide which one to employ in order to estimate X.

Proposition 4. Given $\widehat{X}^{*Y} = V_X W' V_Y^{-1} Y$ and $\widehat{X}^{*X_1} = V_X W_2' V_{X_1}^{-1} X_1$, it is $W_2 \widehat{X}^{*Y} \neq X_1$, and $\widehat{X}^{*Y}$ is less efficient than $\widehat{X}^{*X_1}$ in the sense that it leads to a larger MSDE.

Proof $\widehat{X}^{*Y}$ and $\widehat{X}^{*X_1}$ are obtained by solving the minimization problem in (1) under, respectively, the constraints in (2) and $X_1 = W_2 X$. Hence, it is, $W_2 \widehat{X}^{*Y} = V_{X_1} W_1' V_Y^{-1} W_1 X_1 \neq X_1$. This happens because the constraint $X_1 = W_2 X$ is not considered in the first optimization problem, while (2) is also implicitly imposed in the problem which leads to $\widehat{X}^{*X_1}$.

Then, we have:

$$\begin{gathered} E(X - \widehat{X}^{*Y})(X - \widehat{X}^{*Y})' = E((X - \widehat{X}^{*X_1}) - (\widehat{X}^{*X_1} - \widehat{X}^{*Y}))((X - \widehat{X}^{*X_1}) + \\ -(\widehat{X}^{*X_1} - \widehat{X}^{*Y}))' = E(X - \widehat{X}^{*X_1})(X - \widehat{X}^{*X_1})' + \\ +E(X - \widehat{X}^{*X_1})(\widehat{X}^{*X_1} - \widehat{X}^{*Y})' + \\ E(\widehat{X}^{*X_1} - \widehat{X}^{*Y})(X - \widehat{X}^{*X_1})' + E(\widehat{X}^{*X_1} - \widehat{X}^{*Y})(\widehat{X}^{*X_1} - \widehat{X}^{*Y})'. \end{gathered}$$

But,

$$\begin{gathered}E(X-\widehat{X}^{*X_1})(\widehat{X}^{*X_1}-\widehat{X}^{*Y})'=\\=(I-V_XW_2'V_{X_1}^{-1}W_2)V_XW_2'(V_{X_1}^{-1}-W_1'V_Y^{-1}W_1)W_2V_X=\\V_XW_2'(V_{X_1}^{-1}-W_1'V_Y^{-1}W_1)W_2V_X+\\-V_XW_2'V_{X_1}^{-1}V_{X_1}(V_{X_1}^{-1}-W_1'V_Y^{-1}W_1)W_2V_X=0,\end{gathered}$$

and similarly it can be shown that $E(\widehat{X}^{*X_1}-\widehat{X}^{*Y})(X-\widehat{X}^{*X_1})'=0$.

Hence, it is:

$$\begin{aligned}&E(X-\widehat{X}^{*Y})(X-\widehat{X}^{*Y})'-E(X-\widehat{X}^{*X_1})(X-\widehat{X}^{*X_1})'\\=\;&E(\widehat{X}^{*X_1}-\widehat{X}^{*Y})(\widehat{X}^{*X_1}-\widehat{X}^{*Y})',\end{aligned}$$

so that the term on the left hand side is positive definite. This also implies that

$$tr(E(X-\widehat{X}^{*Y})(X-\widehat{X}^{*Y})')-tr(E(X-\widehat{X}^{*X_1})(X-\widehat{X}^{*X_1})')>0,$$

i.e., $\widehat{X}^{*Y}$ leads to a larger MSDE than $\widehat{X}^{*X_1}$. ■

This result suggests the adoption of the most disaggregated available series for further disaggregation.

3.2 Updating of preliminary estimates

We now wish to modify our procedure to deal with the case where a preliminary estimate of X, Z, is available. As it is often assumed in what we have defined the third disaggregation method in the Introduction, we hypothesise that Z does not satisfy the constraints in (2), namely $Y\neq WZ$. Hence, the problem that we have to solve can be stated as

$$\min_{\widehat{X}} tr(E(X-\widehat{X})(X-\widehat{X})'\mid Z) \tag{8}$$

$$\text{s.t. } Y=WX.$$

The solution is given in the following Proposition.

Proposition 5. The linear minimum MSDE estimator of X is

$$\widehat{X}^* = Z + V_X W' V_Y^{-1}(Y - WZ). \tag{9}$$

Proof If we define $S = X - Z$, $\widehat{S} = \widehat{X} - Z$, and $T = Y - WZ$, the problem in (8) can be reformulated as

$$\min_{\widehat{S}} tr(E(S - \widehat{S})(S - \widehat{S})' \mid Z)$$

$$\text{s.t. } T = WS.$$

It is also $V(S|Z) = V(X)$ and $V(T|Z) = V(Y)$. Thus, from Proposition 1, it follows that $\widehat{S}^* = V_S W' V_T^{-1} T$ and

$$\widehat{X}^* = \widehat{S}^* + Z = Z + V_X W' V_Y^{-1}(Y - WZ). \blacksquare$$

Therefore, $\widehat{X}^*$ in (3) coincides with that in (9) when $Z = 0$.

Denton (1971) proposed to

$$\min_{X} (X - Z)' D (X - Z) \tag{10}$$

$$\text{s.t. } Y = WX,$$

where D is an $Nk \times Nk$ deterministic matrix which depends on a penalty function. The optimal estimator is $\widetilde{X}^* = Z + D^{-1}W'(WD^{-1}W')^{-1}(Y - WZ)$. It follows that $\widehat{X}^*$ in (9) is also a solution for this problem when $D = V_X^{-1}$.

Finally, let us consider a g dimensional disaggregated observable stochastic process $s = \{s_t\}_{t=1}^{\infty}$, group its first Nk realizations into the $Nk \times g$ matrix S, and assume that there exists a static stochastic linear relationship between x_t and s_t. If we call $\widehat{b}$ the *GLS* estimator in a regression of y_t on $W s_t$, then a preliminary estimate of X is $Z = WS\widehat{b}$. For such a choice of Z, (9) is equal to the best linear unbiased estimator of X in Chow and Lin (1971).

3.3 A different objective function

Clements and Hendry (1993) criticize the adoption of mean squared forecast error as a criterion for the evaluation of the forecasting performance of a model, on the grounds that it can be not invariant to isomorphic transformations and does not take into account cross correlation among forecast errors. Their suggestion is to adopt the generalized variance as a criterion, which solves both these problems. Actually, whether these are problems or not can be disputed upon, because an investigator might be interested in a particular specification of the model and willing to ignore cross correlation among forecast errors. However, it seems interesting to study what happens in our framework if we substitute the mean square disaggregation error with the generalized variance, $\det(E(X-\widehat{X})(X-\widehat{X})')$, as the objective function. Thus, we wish to solve the problem:

$$\min_{\widehat{X}} \det(E(X-\widehat{X})(X-\widehat{X})') \qquad \text{s.t.} \quad Y = WX. \tag{11}$$

If we limit the potential solutions to the set $\widehat{X} = PY$, we have

$$\det(E(X-\widehat{X})(X-\widehat{X})') = \det(E(X-PWX)(X-PWX)') = \det((I-PW)V_X(I-PW)') = \det(I-PW)\det(V_X)\det(I-PW)'.$$

Therefore, we can reformulate the problem in (11) as:

$$\min_{P} |\det(I-PW)|, \tag{12}$$

where $|\ |$ indicates the absolute value.

If we define $\widetilde{\omega} = \omega_0 + \omega_1 + ... + \omega_{k-1}$, and

$$\underset{1\times k}{p'} = (1/\widetilde{\omega}, 1/\widetilde{\omega}, ..., 1/\widetilde{\omega}),$$

$$\underset{Nk\times N}{P^*} = \begin{pmatrix} p & 0 & ... & 0 \\ 0 & p & ... & 0 \\ ... & & & \\ 0 & 0 & ... & p \end{pmatrix}$$

then,

Proposition 6. P^* solves the problem in (12).

Proof We have:

$$\underset{Nk\times Nk}{I - P^*W} = \begin{pmatrix} Q & 0 & ... & 0 \\ 0 & Q & ... & 0 \\ ... & & & \\ 0 & 0 & ... & Q \end{pmatrix},$$

$$\underset{k\times k}{Q} = \begin{pmatrix} 1-\omega_0/\widetilde{\omega} & -\omega_1/\widetilde{\omega} & ... & -\omega_{k-1}/\widetilde{\omega} \\ -\omega_0/\widetilde{\omega} & 1-\omega_1/\widetilde{\omega} & ... & -\omega_{k-1}/\widetilde{\omega} \\ ... & & & \\ -\omega_0/\widetilde{\omega} & -\omega_1/\widetilde{\omega} & ... & 1-\omega_{k-1}/\widetilde{\omega} \end{pmatrix}.$$

The first column in Q is equal to the sum of the other $k-1$ columns, with the opposite sign. Hence, $\det(Q) = 0$, and $|\det(I - PW)| = 0$. ■

This leads to, and provides a justification for, very naive disaggregation methods. For example, in the case of a monthly process transformed into a quarterly process by average sampling, an optimal monthly estimate is equal to one third of the corresponding quarterly figure. It is exactly equal to the quarterly figure in the case of point-in-time sampling.

Yet, these values are quite unrealistical, and the fact that the "disaggregator" is completely independent of the original process x is also hard to accept, even if it can be a useful property when no disaggregate information is available. Hence, in the remainder of the paper we stick to the original formulation of the objective function, the MSDE, and to the related best disaggregator.

4 Missing observations, outliers, and forecasting

The procedure in Section 2, possibly extended as in Section 3, can be also easily adapted to provide linear minimum MSDE estimators of missing observations. Actually, this only requires a proper choice of W. For example, if the n_i^{th} observations are missing, $i = 1, ..., m$, W has to be equal to an identity matrix whose n_i^{th} rows have been deleted.

We now explicitly consider the formula of the optimal estimator when the first observation is missing, because it is useful to highlight the relationship with former results.

Proposition 7. If the first observation is missing, the optimal estimator in (3) becomes:

$$\widehat{X}^* = \begin{pmatrix} \underset{1\times Nk-1}{p}^{*} \\ \underset{Nk-1\times Nk-1}{I} \end{pmatrix} Y, \tag{13}$$

where $p^* = v_{1.}V_Y^{-1}$, $v_{1.} = (v_{12}, v_{13}, ..., v_{1Nk})$, $v_{1j} = cov(x_1, x_j)$. Moreover, the variance of the estimator of the missing observation is $v_{11} - v_{1.}V_Y^{-1}v_{1.}'$.

Proof In this case it is

$$\underset{Nk\times Nk-1}{V_X W'} = \begin{pmatrix} v_{12} & v_{13} & \dots & v_{1Nk} \\ v_{22} & v_{23} & \dots & v_{2Nk} \\ \dots & & & \\ v_{Nk2} & v_{Nk3} & \dots & v_{NkNk} \end{pmatrix},$$

and

$$V_Y = \underset{Nk-1\times Nk-1}{W V_X W'} = \begin{pmatrix} v_{22} & v_{23} & \dots & v_{2Nk} \\ v_{32} & v_{33} & \dots & v_{3Nk} \\ \dots & & & \\ v_{Nk2} & v_{Nk3} & \dots & v_{NkNk} \end{pmatrix}.$$

Given that $P^* = V_X W' V_Y^{-1}$, it must be $P^* V_Y = V_X W'$. Let us partition P^*, $P^* V_Y$ and $V_X W'$ as

$$P^* = \begin{pmatrix} \underset{1\times Nk-1}{p} \\ \underset{Nk-1\times Nk-1}{P} \end{pmatrix}, \quad P^* V_Y = \begin{pmatrix} pV_Y \\ PV_Y \end{pmatrix}, \quad V_X W' = \begin{pmatrix} v_{1.} \\ V_Y \end{pmatrix}.$$

It follows that it must be $P = I$ and $p = v_{1.}V_Y^{-1}$. This also implies that the variance of the estimator of the missing observation is $v_{11} - v_{1.}V_Y^{-1}v_{.1}'$. ■

Hence, the best estimators of the available observations are the observations themselves, as one could have expected, while that of the missing observation is a linear combination of the available ones. A more familiar interpretation of the weights is possible. Actually,

Proposition 8. The i^{th} element of p^* is equal to the negative of the i^{th} lag of the inverse autocorrelation function, ρ_i, of x, $i = 1, ..., Nk - 1$.

Proof Kato (1984) showed that the inverse autocorrelation function at lag $j$is equal to the negative of the partial correlation between x_t and x_{t-j}. From Brockwell and Davis (1991, p. 102), the partial correlations between x_t and x_{t-j}, $j = 1, ..., Nk - 1$, can be written, in our notation, as $v_{1.}V_Y^{-1}$. ■

The inverse autocorrelation function is the basic tool in the literature on the estimation of missing observations. In particular, under the additional hypothesis that x is a doubly infinite process, it is well known that the minimum MSDE estimator of the missing observation x_i is:

$$\widehat{x}_i^* = -\sum_{j=1}^{\infty} \rho_j (x_{i-j} + x_{i+j}), \tag{14}$$

see, e.g., Whittle (1963). (14) can be seen as the limit of

$$\widehat{x}_i^* = - \sum_{j=1}^{(Nk-1)/2} \rho_j (x_{i-j} + x_{i+j}), \tag{15}$$

when N tends to infinity, where (15) is our optimal estimator for the mid-sample observation.

We now hypothesize that the series has additive outliers, w_i, in periods $i = n_1, ..., n_m$, so that the observed values are $y_i = x_i + w_i$. The optimal linear estimators of the outliers, $\widehat{w}_i^*$, can be constructed by subtracting from y_i the optimal estimators of x_i, when x_i are treated as missing observations, i.e.,

$$\widehat{w}_i^* = y_i - \widehat{x}_i^*. \tag{16}$$

Their variance covariance matrix is equal to that of the estimators of the missing observations. This approach was followed, e.g., in Peña and

Maravall (1991) just to relate the estimation of outliers and missing observations.

As an example, if there is one outlier in the middle of the sample, its best linear estimator according to our procedure is:

$$\widehat{w}_i^* = y_i + \sum_{j=1}^{(Nk-1)/2} \rho_j (x_{i-j} + x_{i+j}). \tag{17}$$

The limit of this expression when N diverges coincides with the formula for the optimal estimator of an additive outlier in Chang *et al.* (1988).

Finally, notice that forecasting can be considered as a problem of estimation of a set of missing observations at the end of the sample. Hence, linear minimum MSFE forecasts, and their associated standard errors, can be also easily obtained by means of a proper choice of W.

5 Identification of $\mathbf{V}_x$

So far we have hypothesised that V_x or the form of the original process are known. Instead, we now investigate whether and how available aggregate information can be used to infer them. In the next Section we examine the links between the aggregated and disaggregated generating processes, while in this Section we analyse in greater details the relationship between V_x and V_y. In particular, we have noticed that

$$WV_XW' = V_Y, \tag{18}$$

because of the deterministic relationship $Y = WX$. We now consider under what conditions V_X can be uniquely determined given V_Y and W.

It is useful to partition X into $X = (X_1, X_2, ..., X_N)'$, where $X_1 = (x_1, x_2, ..., x_k)'$, $X_2 = (x_{k+1}, x_{k+2}, ..., x_{2k})'$, and so on. V_X can then be partitioned into the N^2 $k \times k$ matrices $C_{i,j} = cov(X_i, X_j)$, $i, j = 1, ..., N$. We also define $w = (\omega_0, ..., \omega_{k-1})$ and indicate with $v_{i,j}^X$, $v_{i,j}^Y$, the elements in the ith row and jth column of V_X and V_Y.

It follows that (18) can be rewritten as

$$wC_{i,j}w' = v^Y_{i,j}, \quad i,j = 1,...N. \tag{19}$$

Thus, $v^Y_{i,j}$ is a linear combination of the elements in $C_{i,j}$, whose weights depend on the aggregation scheme.

In (19) there are N^2 equations in N^2k^2 unknowns (the elements of V_X), which can be reduced to $N(N+1)/2$ equations in $Nk(Nk+1)/2$ unknowns because of symmetry of V_X and V_Y. Hence, in general, there are $d = (N^2(k^2-1)+N(k-1))/2$ degrees of freedom in the determination of the elements of V_X given those of V_Y. Thus, a necessary condition to exactly identify V_X, is the imposition of d *a priori* restrictions on its elements, which can be, e.g., in the form of exclusion or linear restrictions.

In the case of weak stationarity of X the number of degrees of freedom is much lower but still high. Actually, (19) can be reduced to

$$wC_{1,j}w' = v^Y_{1,j}, \quad j = 1,...N, \tag{20}$$

because knowledge of $C_{1,1},...,C_{1,N}$ is sufficient to recover V_X. (20) can be also rewritten in the linear system formulation

$$Av^X = v^Y \tag{21}$$

where $v^X = (v^X_{1,1}, v^X_{1,2}, ..., v^X_{1,NK})'$, $v^Y = (v^Y_{1,1}, v^Y_{1,2}, ..., v^Y_{1,N})'$,

$$\underset{N\times NK}{A} = \begin{pmatrix} a_k & 2a_{k-1} & 2a_{k-2} & ... & 2a_1 & 0 & 0 & & & & 0 \\ 0 & a_1 & a_2 & ... & a_{k-1} & a_k & a_{k-1} & ... & 0 & ... & 0 \\ 0 & 0 & a_1 & ... & a_{k-2} & a_{k-1} & a_k & ... & a_1 & ... & 0 \\ ... \\ 0 & 0 & 0 & ... & 0 & 0 & 0 & ... & 0 & ... & a_1 \end{pmatrix}$$

and $a_j = \sum_{i=0}^{j-1} \omega_i \omega_{i+k-j}$, $j = 1,...,k$. A similar expression was derived by Wei and Stram (1990) for the case $w = (1,1,...,1)$.

In (21) there are N equations in Nk unknowns. Therefore, the imposition of $Nk - N$ a priori restrictions on v^X is a necessary condition for its exact identification. A useful example of the non sufficiency of such a condition is that of point-in-time sampling, where $w = (1, 0, ..., 0)$. In this case, it is $v^X_{1,1+kj} = v^Y_{1,1+j}$, $j = 0, ..., N-1$, and the other elements of v^X are completely unrestricted. If they are set equal to fixed numbers v^X is exactly identified, but if some of the $v^X_{1,1+kj}$ are set equal to arbitrary fixed numbers, the system in (21) cannot be solved.

Instead, we have,

Proposition 9. If $Nk - N$ elements of v^X are set equal to fixed numbers, the corresponding columns of A are deleted in order to obtain the $N \times N$ matrix $\widetilde{A}$, and the unrestricted elements of v^X are grouped in the $N \times 1$ vector $\widetilde{v}^X$, sufficient condition for v^X to be exactly identified is that $|\widetilde{A}| \neq 0$, which implies that it is $\widetilde{v}^X = \widetilde{A}^{-1} v^Y$.

Proof The proof is immediate. ■

When a small subset of observations are missing, it is in general possible to recover v^X from v^Y. A relevant exception is when the first or last k observations are missing. In this case, it is not possible to recover the k highest lags of v^X.

In practice, v^Y is also not known and has to be estimated from a set of N realizations. It follows that the estimates of the higher order covariances of Y can be rather inaccurate, and the inaccuracy is transmitted to those of X.

To summarize, given the relationship between V^X and V^Y in (19) or (21), an infinite number of V^Xs is compatible with V^Y in the disaggregation case. When one or more V^Xs are identified by means of a priori restrictions, the methods in the former Sections can be used to obtain the best linear estimator of X. Yet, it can be easier to formulate a priori restrictions on disaggregated processes than directly on the elements of V^X. Thus, we move to studying the relationship between the aggregated and disaggregated generating processes.

6 Identification of the disaggregated process

We now consider the links between the aggregated and disaggregated process, under the assumption that the latter belongs to the ARIMA class. This is a convenient assumption because ARIMA processes seem to adequately statistically characterize many time series, and they can provide accurate approximations also for the V_xs which would be implied by a nonlinear process. There is also a major computational advantage in deriving $\widehat{X}^*$ under the hypothesis that x is an ARIMA process because the order of the matrices V_X and V_Y can be reduced to that of the MA component of the x and y processes, as shown by Anderson (1974). Moreover, the modification in the structure of a non linear process due to temporal aggregation is difficult to handle, being model specific.

In the first subsection we briefly recall some results on temporal aggregation of ARIMA processes, which provides a useful background for the following discussion. In the second subsection we identify the set of ARIMA processes which is compatible with a given aggregate ARIMA process for y. The third subsection presents an example which illustrates the theoretical findings.

6.1 Temporal Aggregation of an ARIMA process

Let us assume that the elements of x satisfy the stochastic difference equations

$$g(L)x_t = f(L)e_t \tag{22}$$

where $g(L)$ and $f(L)$ are polynomials of degree g and f in the lag operator, L, the roots of $f(l) = 0$ are assumed to lie outside the unit circle, and e_t is white noise (WN) with $V(e_t) = \sigma_e$. Given this specification, we are interested in determining the generating mechanism of y.

Such an issue has been dealt with by many authors in the past. The method that we review in this subsection is that proposed by Brewer (1973), refined by Wei (1981) and Weiss (1984), and further extended in Marcellino (1996) to deal with general aggregation schemes, multivariate processes, and determine the coefficients of the MA component. It requires an introduction of a polynomial in the lag operator, $b(L)$, whose degree in L is at most equal to $gk-g$ and which is such that the product $h(L) = b(L)g(L)$ only contains powers of L^k. It can be shown that such a polynomial always exists, and its coefficients depend on those of $g(L)$, see the above references for details.

In order to determine the AR component of the aggregate ARIMA process, we then multiply both sides of (22) by $\omega(L)$ and $b(L)$ to get

$$h(L^k)\omega(L)x_t = f(L)b(L)\omega(L)e_t. \tag{23}$$

The left hand side of (23) can be also written as

$$h(Z)y_\tau,$$

where $Z = L^k$ is the lag operator at the aggregate temporal frequency, i.e., $Zy_\tau = y_{\tau-1}$. Thus, the order of the aggregate AR component, h, is at most equal to g.

Notice that $h(Z)$ can be decomposed into

$$\prod_{i=1}^{h}\prod_{j=1}^{k}(1-\frac{1}{h_{ij}}L), \tag{24}$$

and at least one h_{ij} for each i has to be such that $g(h_{ij}) = 0$.

It can then be shown that the order of the aggregate MA component, q, coincides with the highest multiple of k non zero lag in the autocovariance function of $f(L)b(L)\omega(L)e_t$. Its coefficients have to be such that the implied autocovariances coincide with those of $f(L)b(L)\omega(L)e_t$ evaluated at all multiples of k.

In summary, y evolves according to:

$$h(Z)y_\tau = q(Z)u_\tau, \tag{25}$$

where u_τ is white noise and $V(u_\tau) = \sigma_u$.

6.2 Identification of the disaggregated ARIMA process

Having set up a framework which relates the disaggregated and aggregated ARIMA processes, we now have to determine what and how many ARIMA processes are compatible with an aggregated one. To start with, assuming that y follows the model in (25), we try and identify the $g(L)$ polynomials which can have generated $h(Z)$. This requires to analyse all the possible decompositions of $h(L^k)$ into $b(L)g(L)$.

We have said that at least one h_{ij} for each i in (24) has to be such that $g(h_{ij}) = 0$. The other $k-1$ h_{ij}s can instead solve either $b(h_{ij}) = 0$ or also $g(h_{ij}) = 0$. Thus, for each i, there are $2^k - 1$ possible "distributions" of the h_{ij}s as roots of $b(L)$ and $g(L)$. Hence, we obtain a total of $(2^k-1)^h$ potential disaggregated AR components, which can be written as

$$\prod_m (1 - \frac{1}{h_m} L), \tag{26}$$

where the h_ms are the h_{ij}s which are considered as roots of $g(l) = 0$. The possible degree of $g(L)$ ranges from h to hk.

Let us now consider the disaggregated MA component. Its coefficients have to be such that the autocovariances at lags multiple of k are equal to those of the aggregated MA component, whose first q values are different from zero. Given a decomposition of $h(Z)$ into $b(L)$ and $g(L)$, this condition is usually not sufficient to exactly identify one disaggregated MA component but an infinite number of them is admissible, which mirrors the existence of an infinite number of V_Xs which are compatible with V_Y. As in that situation, it is necessary to resort to a priori restrictions to obtain exact identification.

A simple but rather stringent condition for exact identification of the disaggregate process is:

Proposition 10. All the roots of $g(l) = 0$ are distinct and positive, or distinct and possibly negative if k is even, and $f \leq g$, or $f \leq g - 1$ in the case of point-in-time sampling.

Proof If $g(l) = 0$ has distinct and positive roots, or distinct and possibly negative roots if k is even, then they coincide with those of $h(z) = 0$ raised to power of $1/k$, and this exactly identifies the AR component. Moreover, in this case, the aggregated MA component is of order g, or $g-1$ for point-in-time sampling. Hence, the at most g, or $g-1$, coefficients of the (invertible) disaggregate MA component have to satisfy g, or $g-1$, conditions, and this is possible only if they coincide with those of $f(L)$. ■

Wei and Stram (1990) discuss more general sufficient a priori conditions for one disaggregate model to be identifiable from an aggregate one. As an alternative, many models can be selected on the basis of subjective judgments and the formula in (6) used to determine $\widehat{X}_M^*$.

Finally, it can be worthwhile pointing out that the procedure can be applied even if the variables are integrated of order d. Actually, in this case $X = (x_1, ..., x_{Nk})'$ and $Y = WX$ are non stationary, but they still have finite second order moments. As an alternative, after having determined the best estimators for the stationary d^{th} differences of the variables, $\widehat{dX}^*$, a method which has been proposed by Stram and Wei (1986) can be adopted to recover the disaggregated values in levels.[3]

Actually, they show that it is

$$\widehat{X}^* = \left(\frac{\underset{kn-d\times kn}{\Delta^d}}{\underset{d\times k(n-d)}{0} \;\Big|\; \underset{d\times kd}{I_d \otimes J_k}} \right)^{-1} \begin{pmatrix} \underset{kn-d\times 1}{\widehat{dX}^*} \\ \underset{d\times 1}{Y^*} \end{pmatrix}, \tag{27}$$

where I_d is a $d \times d$ identity matrix, J_k is a $1 \times k$ vector of ones, $\otimes$ denotes the Kronecker product, Y^* is a $d \times 1$ vector which contains the last d elements in Y, and

$$\Delta^d = \begin{pmatrix} d_0 & d_1 & ... & d_d & 0 & ... & & 0 \\ 0 & d_0 & d_1 & ... & d_d & 0 & ... & 0 \\ ... & & & & & & & \\ 0 & ... & & 0 & ... & & d_{d-1} & d_d \end{pmatrix},$$

where d_i is the coefficient of L^i in $(L-1)^d$.

---

[3]We recall that under the assumption of joint normality, our optimal estimator coincides with that in Stram and Wei (1986), even if they are derived by solving different optimization problems, see Section 2.

6.3 An example

We now discuss an example in order to illustrate the identification problem which has been highlighted in the former subsection. A simple ARMA(1,1) disaggregated process whose realizations are subject to average sampling with $k = 2$ is enough to this end.

Hence, we assume that

$$(1 - g_1 L)x_t = (1 + f_1 L)e_t, \quad e_t \sim WN(0,1), \tag{28}$$

and it can be simply verified, by applying the procedure in subsection 6.1, that the aggregate model is still ARMA(1,1). In particular, it is

$$(1 - h_1 Z)y_\tau = (1 + q_1 Z)u_\tau, \quad u_\tau \sim WN(0, \sigma_u), \tag{29}$$

with $h_1 = g_1^2$, $q_1 = c/2 - \sqrt{c^2/4 - 1}$, $c = \gamma_0/\gamma_2$, $\gamma_0 = 1 + (1 + g_1 + f_1)^2 + (g_1 + f_1 + g_1 f_1)^2 + g_1^2 f_1^2$, $\gamma_2 = (1 + g_1 + f_1)g_1 f_1 + (g_1 + f_1 + g_1 f_1)$, $\sigma_u = \gamma_2/f_1$, γ_0 and γ_2 are the variance and autocovariance at lag 2 of $(1 + g_1 L)(1 + L)(1 + f_1 L)e_t$, and we hypothesise $c > 6$.

We now have to consider what disaggregate ARMA models are compatible with that in (29). According to the derivation in subsection 6.2, $(2^k - 1)^h = 3$ potential disaggregated AR components could have generated that in (29). They are:

$$\begin{aligned} &a) \quad (1 - g_1^2 L^2) \text{ which corresponds to } b(L) = 1; \\ &b) \quad (1 - g_1 L) \text{ which corresponds to } b(L) = (1 + g_1 L); \\ &c) \quad (1 + g_1 L) \text{ which corresponds to } b(L) = (1 - g_1 L). \end{aligned} \tag{30}$$

For each of them, we have to identify the set of potential disaggregate MA components. In the case of (30a) the order of the MA component has to be larger than or equal to one. If it is equal to one, i.e.,

$$a1) \quad (1 - g_1^2 L^2)x_t = (1 + f_1^{a1} L)e_t^{a1}, \quad e_t^{a1} \sim WN(0, \sigma_e^{a1}), \tag{31}$$

then, for the covariance equality constraint in subsection 6.2 to hold, it must be $f_1^a = (c-2)/4 - \sqrt{(c-2)^2/16 - 1}$, $\sigma_e^a = \gamma_2/f_1^a$.

If instead we choose the order equal to 2, i.e.,

$$a2) \quad (1-g_1^2L^2)x_t = (1+f_1^{a2}L)(1+f_2^{a2}L^2)e_t^{a2}, \quad e_t^{a2} \sim i.i.d.(0,\sigma_e^{a2}), \quad (32)$$

then, the following constraints have to be valid:

$$[1+(1+f_1^{a2}+f_2^{a2})^2+(f_1^{a2}+f_2^{a2}+f_1^{a2}f_2^{a2})^2+ \\ +(f_1^{a2})^2(f_2^{a2})^2]\sigma_e^{a2}=\gamma_0 \\ [(1+f_1^{a2}+f_2^{a2})f_1^{a2}f_2^{a2}+(f_1^{a2}+f_2^{a2}+f_1^{a2}f_2^{a2})]\sigma_e^{a2}=\gamma_2.$$

This is a system of two equations in three unknowns so that, in general, infinite solutions are admissible. For example, the system can be solved for f_1^{a2} and σ_e^{a2} as functions of f_2^{a2}. In this case, it is obtained:

$$f_1^{a2}=c(1+(f_2^{a2})^2)/2-f_2^{a2}(2+f_2^{a2})+ \\ -\{[c(1+(f_2^{a2})^2)/2-f_2^{a2}(2+f_2^{a2})]^2-(2+2(f_2^{a2})^2+ \\ -f_2^{a2}c)(2+2(f_2^{a2})^2-f_2^{a2}c+2f_2^{a2})\}^{1/2} \\ \sigma_e^{a2}=\gamma_2/(1+f_1^{a2}+f_2^{a2})f_1^{a2}f_2^{a2}+(f_1^{a2}+f_2^{a2}+f_1^{a2}f_2^{a2}).$$

Similarly, if the order of the MA component is assumed to be equal to 3, the constraint in subsection 6.2 determines a system of 3 equations in 4 unknowns, and the indeterminacy persists when increasing the order.

In the case of the AR component (30b), the choice $f=0$ already generates an aggregate ARMA(1,1) model but the covariance equality constraint cannot be satisfied. Hence, it must be $f \geq 1$. When $f=1$, the correct MA component is recovered, while for $f>1$ the indeterminacy problem arises. For the AR component (30c) similar results are obtained, $f=0$ is not admissible, when $f=1$ the MA component can be exactly determined, while for $f>1$ infinite solutions are possible.

The sufficient condition for exact identification in Proposition 9 would lead to the choice of the correct model in this example. This would not be the case if, e.g., the generating model were (31).

In summary, even if the processes are restricted to those in the ARIMA class, an infinite number of them is in general compatible with a given aggregated model for y. This result depends on the under identification of the MA component and, if its exact identification is achieved by means of a priori restrictions, the class of potential disaggregated processes shrinks to contain a finite number of elements. Further a priori restrictions are required to select one element in this class. As an alternative, subjective probabilistic statements about the validity of the restrictions can be expressed, and the formula in (6) used to determine $\widehat{X}_M^*$.

7 Conclusions

In this paper we have suggested a procedure to linearly recover disaggregated values of a time series from the available aggregated values. It minimizes the mean squared disaggregation error and satisfies the compatibility constraint with aggregated values, without requiring assumptions of normality or linearity of the underlying disaggregated stochastic process. When such hypotheses are introduced, the procedure is related to many other disaggregation methods which have been suggested in the literature.

The basic procedure can be extended to estimate many series, possibly at several disaggregated frequencies, or to revise preliminary estimates of disaggregated values. It is so flexible that it can be also adapted to recover missing observations, outliers, linearly disaggregated series, to forecasting, and to deal with uncertainty about the disaggregated process.

The latter issue seems to be important, even if it is rather neglected in the literature. Actually, we have also derived the set of unobservable disaggregated second moments or generating mechanisms which are compatible with their observable aggregated counterparts, and are required for the implementation of disaggregation methods. This set usually contains an infinite number of elements so that particular *a priori* restrictions have to be imposed for the disaggregated characteristics to be exactly identifiable from the aggregated ones. As an alternative, a probabilistic statement can be made on their relative likelihood, and it is incorporated into the disaggregation procedure.

The presence of subjective decisions in the implementation of disaggregation procedures is rather unsatisfactory. Unfortunately, it cannot be eliminated. This suggests that the existing effort to increase the frequency of data collection in order for it to match that of the original process, or at least that of usual interest, is really valuable.

References

[1] Anderson, O.D. (1974), "An Improved Approach to Inverting the Autocovariance Matrix of a General Mixed Autoregressive Moving Average Time Process", *Australian Statistical Journal*, 18, pp. 73-75.

[2] Boot, J.C.G., Feibes, W. and Lisman, J.H.C. (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data", *Applied Statistics*, 16, pp. 65-75.

[3] Brewer, K.R.W. (1973), "Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX models", *Journal of Econometrics*, 1, pp. 133-154.

[4] Brockwell, P.J. and Davis, R.A. (1991), "Time Series: Theory and Methods", Berlin: Springer Verlag.

[5] Chan, W.S. (1993), "Disaggregation of Annual Time-Series Data into Quarterly Figures: A Comparative Study", *Journal of Forecasting*, 12, pp. 677-688.

[6] Chang, I., Tiao, G.C. and Chen, Ch. (1988), "Estimation of Time Series Parameters in the Presence of Outliers", *Technometrics*, 30, pp. 193-204.

[7] Chow, G.C. and Lin, A. (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time-Series by Related Series", *Review of Economics and Statistics*, 53, pp. 372-375.

[8] Clements, M.P. and Hendry, D.F. (1993), "On the Limitations of Comparing Mean Squared Forecast Errors", *Journal of Forecasting*, 12, pp. 617-637.

[9] Cohen, K.J., Müller, W. and Padberg, M.W. (1971), "Autoregressive Approaches to Disaggregation of Time Series Data", *Applied Statistics*, 20, pp. 119-129.

[10] Denton, F. (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization", *Journal of the American Statistical Association*, 66, pp. 99-101.

[11] Di Fonzo, T. (1990), "The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates Are Known", *Review of Economics and Statistics*, pp. 178-182.

[12] Fernandez, R.B. (1981), "A Methodological Note on the Estimation of Time-Series", *Review of Economics and Statistics*, 63, pp. 471-476.

[13] Gomez, V. and Maravall, A. (1994), "Estimation, Prediction and Interpolation for Non Stationary Time Series with the Kalman Filter", *Journal of the American Statistical Association*, 89, pp. 611-624.

[14] Guerrero, V.M. (1990), "Temporal Disaggregation of Time-Series: An ARIMA-Based Approach", *International Statistical Review*, 58, pp. 29-46.

[15] Harvey, A.C. and Pierse, R.G. (1984), "Estimating Missing Observations in Economic Time Series", *Journal of the American Statistical Association*, 79, pp. 125-131.

[16] Kato, A.J. (1984), "A Characterization of the Inverse Autocorrelation Function", *Communications in Statistics, Theory and Methods*, 13, pp. 2503-2510.

[17] Kohn, R. and Ansley, C.F. (1986), "Estimation, Prediction and Interpolation for ARIMA models with Missing Data", *Journal of the American Statistical Association*, 81, pp. 751-761.

[18] Lisman J.H.C. and Sandee, J. (1964), "Derivation of Quarterly Figures from Annual Data", *Applied Statistics*, 13, pp. 87-90.

[19] Litterman, R.B. (1983), "A Random Walk, Markov Model for the Distribution of Time Series", *Journal of Business and Economic Statistics*, 1, pp. 169-173.

[20] Marcellino, M. (1996), "Some Temporal Aggregation Issues in Empirical Analysis", W.P. UCSD 96/39.

[21] Nijman, T.E. and Palm, F.C. (1986), The Construction and Use of Approximations for Missing Quarterly Observations: A Model-Based Approach", *Journal of Business and Economic Statistics*, 4, pp. 47-58.

[22] Pena, D. and Maravall, A. (1991), "Interpolation, Outliers and Inverse Auotcorrelations", *Communications in Statistics, Theory and Methods*, 20, pp. 3175-3186.

[23] Rossi, N. (1982), "A Note on the Estimation of Disaggregate Time Series when the Aggregate is Known", *Review of Economics and Statistics*, pp. 695-696.

[24] Stram, D.O. and Wei, W.W.S. (1986), "A Methodological Note on Disaggregation of Time Series Totals", *Journal of Time Series Analysis*, 7, pp. 293-302.

[25] Wei, W.W.S. (1981), "Effects of systematic sampling on ARIMA models", *Communications in Statistical-Theoretical Mathematics*, 10, pp. 2389-2398.

[26] Wei, W.W.S. and Stram, D.O. (1990), "Disaggregation of Time Series Models", *Journal of the Royal Statistical Society*, Series B, 52, pp. 453-467.

[27] Weiss, A.A. (1984), "Systematic Sampling and Temporal Aggregation in Time Series Models", *Journal of Econometrics*, pp. 271-281.

[28] Whittle, P. (1963), "Prediction and Interpolation by Linear Least Squares Methods", London: English Univerisities Press.

EUI WORKING PAPERS

EUI Working Papers are published and distributed by the European University Institute, Florence

Copies can be obtained free of charge – depending on the availability of stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

To The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI) – Italy
Telefax No: +39/55/4685 636
e-mail: publish@datacomm.iue.it
http://www.iue.it

From Name . . . .
Address. . . . .
. . . .
. . . .
. . . .
. . . .

❒ Please send me a complete list of EUI Working Papers
❒ Please send me a complete list of EUI book publications
❒ Please send me the EUI brochure Academic Year 1998/99

Please send me the following EUI Working Paper(s):

No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .
No, Author . . . .
Title: . . . .

Date . . . .

Signature . . . .

Working Papers of the Department of Economics Published since 1996

ECO No. 96/1
Ana Rute CARDOSO
Earnings Inequality in Portugal: High and Rising?

ECO No. 96/2
Ana Rute CARDOSO
Workers or Employers: Who is Shaping Wage Inequality?

ECO No. 96/3
David F. HENDRY/Grayham E. MIZON
The Influence of A.W.H. Phillips on Econometrics

ECO No. 96/4
Andrzej BANIAK
The Multimarket Labour-Managed Firm and the Effects of Devaluation

ECO No. 96/5
Luca ANDERLINI/Hamid SABOURIAN
The Evolution of Algorithmic Learning: A Global Stability Result

ECO No. 96/6
James DOW
Arbitrage, Hedging, and Financial Innovation

ECO No. 96/7
Marion KOHLER
Coalitions in International Monetary Policy Games

ECO No. 96/8
John MICKLEWRIGHT/ Gyula NAGY
A Follow-Up Survey of Unemployment Insurance Exhausters in Hungary

ECO No. 96/9
Alastair McAULEY/John MICKLEWRIGHT/Aline COUDOUEL
Transfers and Exchange Between Households in Central Asia

ECO No. 96/10
Christian BELZIL/Xuelin ZHANG
Young Children and the Search Costs of Unemployed Females

ECO No. 96/11
Christian BELZIL
Contiguous Duration Dependence and Nonstationarity in Job Search: Some Reduced-Form Estimates

ECO No. 96/12
Ramon MARIMON
Learning from Learning in Economics

ECO No. 96/13
Luisa ZANFORLIN
Technological Diffusion, Learning and Economic Performance: An Empirical Investigation on an Extended Set of Countries

ECO No. 96/14
Humberto LÓPEZ/Eva ORTEGA/Angel UBIDE
Explaining the Dynamics of Spanish Unemployment

ECO No. 96/15
Spyros VASSILAKIS
Accelerating New Product Development by Overcoming Complexity Constraints

ECO No. 96/16
Andrew LEWIS
On Technological Differences in Oligopolistic Industries

ECO No. 96/17
Christian BELZIL
Employment Reallocation, Wages and the Allocation of Workers Between Expanding and Declining Firms

ECO No. 96/18
Christian BELZIL/Xuelin ZHANG
Unemployment, Search and the Gender Wage Gap: A Structural Model

ECO No. 96/19
Christian BELZIL
The Dynamics of Female Time Allocation upon a First Birth

ECO No. 96/20
Hans-Theo NORMANN
Endogenous Timing in a Duopoly Model with Incomplete Information

*out of print

ECO No. 96/21
Ramon MARIMON/Fabrizio ZILIBOTTI
'Actual' Versus 'Virtual' Employment in Europe: Is Spain Different?

ECO No. 96/22
Chiara MONFARDINI
Estimating Stochastic Volatility Models Through Indirect Inference

ECO No. 96/23
Luisa ZANFORLIN
Technological Diffusion, Learning and Growth: An Empirical Investigation of a Set of Developing Countries

ECO No. 96/24
Luisa ZANFORLIN
Technological Assimilation, Trade Patterns and Growth: An Empirical Investigation of a Set of Developing Countries

ECO No. 96/25
Giampiero M.GALLO/Massimiliano MARCELLINO
In Plato's Cave: Sharpening the Shadows of Monetary Announcements

ECO No. 96/26
Dimitrios SIDERIS
The Wage-Price Spiral in Greece: An Application of the LSE Methodology in Systems of Nonstationary Variables

ECO No. 96/27
Andrei SAVKOV
The Optimal Sequence of Privatization in Transitional Economies

ECO No. 96/28
Jacob LUNDQUIST/Dorte VERNER
Optimal Allocation of Foreign Debt Solved by a Multivariate GARCH Model Applied to Danish Data

ECO No. 96/29
Dorte VERNER
The Brazilian Growth Experience in the Light of Old and New Growth Theories

ECO No. 96/30
Steffen HÖRNIG/Andrea LOFARO/Louis PHLIPS
How Much to Collude Without Being Detected

ECO No. 96/31
Angel J. UBIDE
The International Transmission of Shocks in an Imperfectly Competitive International Business Cycle Model

ECO No. 96/32
Humberto LOPEZ/Angel J. UBIDE
Demand, Supply, and Animal Spirits

ECO No. 96/33
Andrea LOFARO
On the Efficiency of Bertrand and Cournot Competition with Incomplete Information

ECO No. 96/34
Anindya BANERJEE/David F. HENDRY/Grayham E. MIZON
The Econometric Analysis of Economic Policy

ECO No. 96/35
Christian SCHLUTER
On the Non-Stationarity of German Income Mobility (and Some Observations on Poverty Dynamics)

ECO No. 96/36
Jian-Ming ZHOU
Proposals for Land Consolidation and Expansion in Japan

ECO No. 96/37
Susana GARCIA CERVERO
Skill Differentials in the Long and in the Short Run. A 4-Digit SIC Level U.S. Manufacturing Study

***

ECO No. 97/1
Jonathan SIMON
The Expected Value of Lotto when not all Numbers are Equal

ECO No. 97/2
Bernhard WINKLER
Of Sticks and Carrots: Incentives and the Maastricht Road to EMU

ECO No. 97/3
James DOW/Rohit RAHI
Informed Trading, Investment, and Welfare

*out of print

ECO No. 97/4
Sandrine LABORY
Signalling Aspects of Managers' Incentives

ECO No. 97/5
Humberto LÓPEZ/Eva ORTEGA/Angel UBIDE
Dating and Forecasting the Spanish Business Cycle

ECO No. 97/6
Yadira GONZÁLEZ de LARA
Changes in Information and Optimal Debt Contracts: The Sea Loan

ECO No. 97/7
Sandrine LABORY
Organisational Dimensions of Innovation

ECO No. 97/8
Sandrine LABORY
Firm Structure and Market Structure: A Case Study of the Car Industry

ECO No. 97/9
Elena BARDASI/Chiara MONFARDINI
The Choice of the Working Sector in Italy: A Trivariate Probit Analysis

ECO No. 97/10
Bernhard WINKLER
Coordinating European Monetary Union

ECO No. 97/11
Alessandra PELLONI/Robert WALDMANN
Stability Properties in a Growth Model

ECO No. 97/12
Alessandra PELLONI/Robert WALDMANN
Can Waste Improve Welfare?

ECO No. 97/13
Christian DUSTMANN/Arthur van SOEST
Public and Private Sector Wages of Male Workers in Germany

ECO No. 97/14
Søren JOHANSEN
Mathematical and Statistical Modelling of Cointegration

ECO No. 97/15
Tom ENGSTED/Søren JOHANSEN
Granger's Representation Theorem and Multicointegration

ECO No. 97/16
Søren JOHANSEN/
Ernst SCHAUMBURG
Likelihood Analysis of Seasonal Cointegration

ECO No. 97/17
Maozu LU/Grayham E. MIZON
Mutual Encompassing and Model Equivalence

ECO No. 97/18
Dimitrios SIDERIS
Multilateral Versus Bilateral Testing for Long Run Purchasing Power Parity: A Cointegration Analysis for the Greek Drachma

ECO No. 97/19
Bruno VERSAEVEL
Production and Organizational Capabilities

ECO No. 97/20
Chiara MONFARDINI
An Application of Cox's Non-Nested Test to Trinomial Logit and Probit Models

ECO No. 97/21
James DOW/Rohit RAHI
Should Speculators be Taxed?

ECO No. 97/22
Kitty STEWART
Are Intergovernmental Transfers in Russia Equalizing?

ECO No. 97/23
Paolo VITALE
Speculative Noise Trading and Manipulation in the Foreign Exchange Market

ECO No. 97/24
Günther REHME
Economic Growth, (Re-)Distributive Policies, Capital Mobility and Tax Competition in Open Economies

*out of print

ECO No. 97/25
Susana GARCIA CERVERO
A Historical Approach to American Skill Differentials

ECO No. 97/26
Susana GARCIA CERVERO
Growth, Technology and Inequality: An Industrial Approach

ECO No. 97/27
Bauke VISSER
Organizational Structure and Performance

ECO No. 97/28
Pompeo DELLA POSTA
Central Bank Independence and Public Debt Convergence in an Open Economy Dynamic Game

ECO No. 97/29
Matthias BRUECKNER
Voting and Decisions in the ECB

ECO No. 97/30
Massimiliano MARCELLINO
Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model Based Procedure

*out of print